Cavity Method for Supersymmetry Breaking Spin Glasses

Andrea Cavagna[†], Irene Giardina[†] and Giorgio Parisi[†‡]
[†] *Dipartimento di Fisica, Università di Roma "La Sapienza" and Center for Statistical Mechanics and Complexity, INFM Roma 1 P.le Aldo Moro 2, 00185 Roma, Italy*
[‡] *Istituto Nazionale di Fisica Nucleare, Sezione Roma 1*

(Dated: July 16, 2004)

The spontaneous supersymmetry-breaking that takes place in certain spin-glass models signals a particular fragility in the structure of metastable states of such systems. This fragility is due to the presence of at least one marginal mode in the Hessian of the free energy, that makes the states highly susceptible under external perturbations. The cavity method is a technique that recursively describes the property of a system with $N+1$ spins in terms of those of a system with N spins. To do so, the cavity method assumes a certain degree of stability when adding a new spin to the system, i.e. it assumes that for a generic choice of the parameters there is an one-to-one correspondence between the metastable states of the system with N spins and the metastable states of the system with $N+1$ spins. In systems where the supersymmetry is broken such a correspondence does not exist, and an alternative formulation of the cavity method must be devised. We introduce a generalized cavity approach that takes care of this problem and we apply it to the computation of the probability distribution of the local magnetizations in the Sherrington-Kirkpatrick model. Our findings agree with the correct supersymmetry-breaking result.

In physics we often meet two quite different situations. On the one hand, we tend to assume (or hope) that small perturbations have small effects. This is at the basis of all attempts to use any sort of perturbation theory. On the other hand, we know that many interesting phenomena take place when such an assumption is in fact violated. At criticality, whenever the susceptibility of a system is infinite, its response is anomalous. In this case small perturbations may indeed have big effects. This state of affairs is often accompanied by the presence of zero, or marginal, modes among the second derivatives of the action, may these be the masses, or the inverse susceptibilities. When this happens we must be extremely careful in applying perturbation theory.

A particularly interesting realization of this scenario occurs when criticality is induced by spontaneous symmetry breaking. In that case, symmetry breaking, marginality, and break-down of standard perturbative techniques, give rise to some interesting phenomena. In this work we discuss how the spontaneous breaking of a supersymmetry is connected to the presence of marginal modes and anomalous response in certain spin-glass models. The main focus of our study is the structure of metastable states in such models, and in particular the extreme fragility of this structure under small perturbations. A significant consequence of this feature is the break-down of the standard cavity method [1, 2], whose basic assumption is that by adding one extra degree of freedom to a large systems, its physical properties (and in particular some properties of the structure of states) do not change dramatically. This assumption is no longer valid when the supersymmetry is broken. Our main result is to provide a generalization of the cavity method that works also in supersymmetry-breaking systems.

Metastable states in mean-field spin-glasses can be identified with the local minima of a mean-field free energy F (also known as the TAP free energy [3]), that is a function of the local magnetizations m_i of the system. Generally speaking, in order to compute the number of local minima of a function $F(\mathbf{m})$, one can introduce an effective action that is invariant under a generalized form of the Becchi-Rouet-Stora-Tyutin (BRST) supersymmetry [4, 5, 6, 7, 8]. The Ward identities generated by this symmetry have a rather clear physical meaning. The most relevant one reads [8],

$$\left.\frac{d\langle m_i\rangle}{dh_j}\right|_{h=0} = \langle A_{ij}(\mathbf{m})\rangle \ , \tag{1}$$

where $A_{ij}(\mathbf{m}) = [\partial_i\partial_j F(\mathbf{m})]^{-1}$ is the inverse of the second derivative, i.e. of the Hessian. The brackets $\langle\cdot\rangle$ indicate a sum over all metastable states of the system. The meaning of equation (1) is straightforward: it expresses the natural relationship between the susceptibility and the curvature of the minima of the free energy. Therefore, the supersymmetry seems to encode a very robust physical feature of metastable states, namely the static fluctuation-dissipation theorem. This notwithstanding, it is now believed that in certain spin-glass models the supersymmetry is in fact spontaneously broken in the low-temperature phase, and that relation (1) is thus violated [9, 10, 11]. Such systems include the Sherrington-Kirkpatrick (SK) model [12], that we discuss here.

The reason why the static fluctuation-dissipation theorem is violated, and thus the supersymmetry is broken, lies in the peculiar structure of metastable states of such systems. Recent studies show [9, 10, 11] that at low temperatures

all stationary points of the mean-field free energy are organized into minimum-saddle pairs. The minimum and the saddle are connected along a mode that is softer the larger the system size N. Moreover, the free energy difference of the paired stationary points decreases with increasing N. In other words, supersymmetry-breaking metastable states are marginal in the thermodynamic limit, having at least one flat zero mode. In this situation it is clear that even an infinitesimal external field may destabilize some states, making them disappear. On the other hand, virtual states, i.e. inflection points of the free energy with a very small second derivative, may be stabilized by the field, giving rise to pairs of new states. In such a situation we must reconsider the validity of equation (1). At the l.h.s. we differentiate with respect to an external field a sum over all metastable states. The problem is that, due to marginality, *some elements in this sum may disappear or appear as the field goes to zero.* Therefore, even though the static fluctuation-dissipation relation holds for each individual state, when we sum over all states, an anomalous contribution arises due to the instability of the whole structure with respect to the field. Relation (1) is thus violated.

Supersymmetry breaking is thus the mathematical expression of a great instability in the structure of metastable states. Adding a new spin to the system, the basic building block of the cavity method, will therefore not be harmless. Our aim is to find a more general formulation of the method, valid also in supersymmetry-breaking systems. Note that a similar task has been independently pursued in [13], although introducing an explicit violation of the supersymmetry. Our starting point is the set of equations satisfied by the local magnetizations in a system with N degrees of freedom. These are just the equations of stationarity of the free energy,

$$\frac{\partial F}{\partial m_i} \equiv E_i^{(N)}(\,\mathbf{m}) = 0 \qquad i = 1, \ldots, N \;, \tag{2}$$

where $\mathbf{m} = m_1, \ldots, m_N$. For the SK model we have [3],

$$E_i^{(N)} \;\equiv\; \tanh^{-1} m_i + \beta^2(1-q)\, m_i - \beta \sum_{j=1}^{N} J_{ij} m_j \;. \tag{3}$$

The self-overlap is $q = (1/N)\sum_i m_i^2$, and the random couplings J_{ij} are drawn from a Gaussian distribution of mean zero and variance $1/N$. Let us introduce the number $d\mathcal{N}^{(N)}(\mathbf{m})$ of solutions of eqs.(2) with magnetizations between $\mathbf{m}$ and $\mathbf{m} + d\mathbf{m}$. The density of solutions $\rho^{(N)}(\mathbf{m})$ is then defined as,

$$d\mathcal{N}^{(N)}(\mathbf{m}) = \rho^{(N)}(\mathbf{m})\; d\mathbf{m} \;. \tag{4}$$

The basic point of the cavity method [1, 2] is to write a recursive relation expressing the density $\rho^{(N+1)}$ of the system with $N+1$ spin, as a function of the density $\rho^{(N)}$ of the system with N spins. The fundamental idea behind this procedure is that adding one single spin to a system with a large number of degrees of freedom, is a small perturbation and consequently the properties of the new $N+1$ system will not be much different from the original N system. More specifically, a natural assumption is that the structure of metastable states of the system does not change when we add the new spin: the solutions of (2) are slightly modified, but there is still a one-to-one correspondence between solutions of the N and $N+1$ systems. This hypothesis seems reasonable, and it has been at the basis of all standard cavity investigations. However, we will show that it is in fact equivalent to assume that the supersymmetry is unbroken.

Let us add a new spin at site 0, and call $(\mathbf{m}, m_0)$ the global magnetization vector of the $N+1$ system. The new set of $N+1$ equations can be split as,

$$E_i^{(N+1)}(\mathbf{m}, m_0) \;=\; 0 \qquad i = 1, \ldots, N \;, \tag{5}$$

$$E_0(\mathbf{m}, m_0) \;=\; 0 \;. \tag{6}$$

We can examine this set of equations by first solving the first N equations (5) at fixed m_0, and then plugging $\mathbf{m}(m_0)$ into (6). The assumption of stability of states can be expressed by saying that there is a one-to-one mapping between solutions $\mathbf{m}^{(N+1)}$ of eqs.(5) and solutions $\mathbf{m}^{(N)}$ of eqs.(2): if $\mathbf{m}^{(N)}$ is a solution of the N system, the effect of adding the spin is just to slightly modify the old magnetization to a new value $\mathbf{m}^{(N+1)} = \mathbf{f}_{m_0}(\mathbf{m}^{(N)})$. From this key hypothesis, it follows that for a given m_0 the number of solutions with magnetization $\mathbf{m}^{(N+1)}$ of the first N equations of the $N+1$ system, is simply equal to the number of solutions of the original N system that have magnetization $\mathbf{m}^{(N)}$,

$$d\mathcal{N}^{(N+1)}(\mathbf{m}^{(N+1)}) = d\mathcal{N}^{(N)}(\mathbf{m}^{(N)}) \;, \tag{7}$$

with $\mathbf{m}^{(N+1)} = \mathbf{f}_{m_0}(\mathbf{m}^{(N)})$. In order to obtain the density of solutions of the new $N+1$ system we must also impose the extra equation (6) for m_0. This relation can be cast in a simple physical form, namely $m_0 = \tanh\left(\beta \sum_j J_{0j} m_j^{(N)}\right)$:

the spin at zero aligns to the local field produced on site 0 by the original N system. The density of solutions for the new system therefore becomes, (sums over repeated indices are understood),

$$\rho^{(N+1)}\left(\mathbf{m}^{(N+1)}, m_0\right)\, d\mathbf{m}^{(N+1)}\, dm_0 = \rho^{(N)}\left(\mathbf{m}^{(N)}\right)\, d\mathbf{m}^{(N)}\;\; \delta\left(\tanh^{-1} m_0 - \beta J_{0j} m_j^{(N)}\right)\;\frac{1}{1-m_0^2}\, dm_0\ , \tag{8}$$

The first part of this equation is equivalent to relation (7), while the second part makes sure that the equation for m_0 is satisfied (the last factor is the Jacobian of the δ-function). The meaning of this equation is clear: the density of solutions remains the same, provided that we satisfy the equation for the new site 0, and that we change a little the other N magnetizations. The map between the old and the new magnetizations can be found by noting that the new magnetization m_0 exerts a field $J_{i0}m_0$ on the old sites, and thus $m_i^{(N+1)} = m_i^{(N)} + \chi\, J_{i0}\, m_0$, where $\chi = \beta(1-q)$ is the susceptibility.

Equation (8) is a recursive relation for the number of states. In order to get a self-consistency equation, we have to work with the probability, i.e. the density divided by the total number of states, $\rho(\mathbf{m})/\mathcal{N}$, that has a well defined limit when $N \to \infty$. In the annealed approximation we assume that the probability is self-averaging and that it can therefore be averaged over the disorder [14]. The averaged probability density factorizes, i.e. $\overline{\rho(\mathbf{m})/\mathcal{N}} = \prod_i p(m_i)$, and thus Eq. (8) becomes a self-consistency equation for the single site probability $p(m_i)$: we divide both sides by the global number of states $\mathcal{N}^{(N+1)}$, average over the disorder, and integrate over the first N magnetizations. In the thermodynamic limit $p^{(N+1)} = p^{(N)}$, and we finally get,

$$p(m_0) = \int \prod_i dP(J_{i0})\; dm_i\; p(m_i)\; \mathcal{K}_{\rm ss}(\mathbf{m}, m_0)\ , \tag{9}$$

where,

$$\mathcal{K}_{\rm ss}(\mathbf{m}, m_0) = \delta\left(\tanh^{-1} m_0 - \beta J_{0j} m_j\right)/(1-m_0^2)\ . \tag{10}$$

The factor $\mathcal{N}^{(N)}/\mathcal{N}^{(N+1)}$ on the r.h.s. has been reabsorbed into the normalization constant of the probability p. This equation is not difficult to solve, and we find,

$$p(m_0) = \frac{1}{\sqrt{2\pi\beta^2 q}} \exp\left[-\frac{(\tanh^{-1} m_0)^2}{2\beta^2 q}\right]\;\frac{1}{1-m_0^2}\ , \tag{11}$$

where $q = \langle m^2 \rangle$, and where $\langle X \rangle = \int dm\, p(m)\, X(m)$. Equation (11) coincides with the probability distribution found in [7], that is the *supersymmetric* distribution. The assumption of stability of metastable states is thus equivalent to assuming unbroken supersymmetry. However, the BRST supersymmetry is in fact broken in the SK model [9, 10], and thus the correct distribution is not given by (11). To find the supersymmetry-breaking distribution within the cavity approach, we must give up the assumption of stability of states, and find a more general formulation. The first step is to recognize that eqs.(5) for the $N+1$ system can be formally rewritten as,

$$E_i^{(N+1)}(\mathbf{m}, m_0|\beta) = E_i^{(N)}(\mathbf{m}|\beta') - k_i(\mathbf{m}, m_0) = 0 \tag{12}$$

with,

$$k_i(\mathbf{m}, m_0) = \beta J_{i0} m_0 - \frac{\beta^2}{N}(1-m_0^2)\, m_i\ . \tag{13}$$

The variance of J_{ij} in the $N+1$ system is smaller than in the N system. We have then rescaled the inverse temperature from β to $\beta' = \beta[1-1/(2N)]$ in such a way that the disorder appearing in $E_i^{(N)}$ has the correct scaling for a system of size N. Equation (12) suggests the path we have to follow: the function k_i may be seen as a local field acting on site i of the N system, so that, in a way, finding solutions of the $N+1$ system is like finding solutions of the N system with a field. If this can be done, we may hope to write a self-consistency equation not simply for the probability density of solutions with given magnetization, but for a more complicated object, i.e. the probability density of solutions with given magnetization *and* field. The equations for the metastable states of an N system with a field are,

$$E_i^{(N)}(\mathbf{m}) = h_i\ . \tag{14}$$

We define the new density of solutions $\rho^{(N)}(\mathbf{m}|\mathbf{h})$ as,

$$d\mathcal{N}^{(N)}(\mathbf{m}|\mathbf{h}) = \rho^{(N)}(\mathbf{m}|\mathbf{h})\; d\mathbf{m}\ , \tag{15}$$

where $d\mathcal{N}^{(N)}(\mathbf{m}|\mathbf{h})$ is the number of solutions of (14), with given magnetization $\mathbf{m}$ and *external* field $\mathbf{h}$. The idea is to write a recursive relation for $\rho^{(N)}(\mathbf{m}|\mathbf{h})$ rather than for $\rho^{(N)}(\mathbf{m})$. From (12) we have that the equations for the $N+1$ system in an external field $(\mathbf{h}, h_0)$ are,

$$E_i^{(N)}(\mathbf{m}) = k_i(\mathbf{m}, m_0) + h_i \tag{16}$$

$$E_0(\mathbf{m}, m_0) = h_0 \ . \tag{17}$$

Let us focus now on (16), at fixed value of m_0. The number of solutions of this set of equations seems formally equal to the number of solutions of a system with external field $\mathbf{k}+\mathbf{h}$. The only problem is that $\mathbf{k}$ is *not* an external field, since it depends on $\mathbf{m}$, and thus $d\mathcal{N}(\mathbf{m}|\mathbf{h}+\mathbf{k}(\mathbf{m})) \neq \rho(\mathbf{m}|\mathbf{h}+\mathbf{k}(\mathbf{m}))\, d\mathbf{m}$. Fortunately, it can be simply proved a similar relation, that reads,

$$d\mathcal{N}^{(N)}(\mathbf{m}|\mathbf{h}+\mathbf{k}(\mathbf{m})) = \rho^{(N)}(\mathbf{m}|\mathbf{h}+\mathbf{k}(\mathbf{m}))\ \omega(\mathbf{m})\ d\mathbf{m} \ , \tag{18}$$

where the re-weighting factor is given by,

$$\omega(\mathbf{m}) = \frac{\partial_{\mathbf{m}}[E_i^{(N)}(\mathbf{m}) - k_i(\mathbf{m})]}{\partial_{\mathbf{m}}[E_i^{(N)}(\mathbf{m})]} \ . \tag{19}$$

We can apply this general formula to equation (16), and remember that (as in the previous case) to obtain the full solutions density for the $N+1$ system we must also satisfy the extra equation for m_0. This time, however, there is no mapping between solutions of the N and $N+1$ systems, and thus the field experienced by spin 0 must be expressed as a function of the new magnetizations [2],

$$E_0 = \tanh^{-1} m_0 - \beta \sum_j J_{0j} m_j + \beta^2 (1-q)\, m_0 \ , \tag{20}$$

and we have to remember that $\mathbf{m}$ depends on m_0. In conclusion, we have,

$$\begin{aligned} \rho^{(N+1)}(\mathbf{m}, m_0|\mathbf{h}, h_0) &= \rho^{(N)}(\mathbf{m}|\mathbf{h}+\mathbf{k}(\mathbf{m}, m_0))\ \omega(\mathbf{m}, m_0) \times \\ &\times\ \delta\left(E_0(\mathbf{m}, m_0) - h_0\right)\ \frac{dE_0}{dm_0} \ . \end{aligned} \tag{21}$$

Equation (21) is a general recursive relation for the density of states, that is *always* valid, whether the structure of states is stable or not. The supersymmetric result can be recovered if we assume that there is a mapping between solutions $\mathbf{m}$ of the $N+1$ system and solutions $\mathbf{f}_{m_0}^{-1}(\mathbf{m})$ of the system in absence of the 0 site. In this case we have (for any fixed N),

$$\rho(\mathbf{m}|\mathbf{h}+\mathbf{k}(\mathbf{m}, m_0))\ \omega(\mathbf{m}, m_0)\ d\mathbf{m} = \rho(\mathbf{f}_{m_0}^{-1}(\mathbf{m})|\mathbf{h})\ d\mathbf{f}_{m_0}^{-1}. \tag{22}$$

In other words, in the supersymmetric case the effective field $\mathbf{k}$ can be reabsorbed via a change of variables, giving back the supersymmetric expression (8).

When we pass from single instances to probabilities, we can average equation (21) over the disorder and once again write self-consistency equation for the single-site probability $p(m_i|h_i)$. After some algebra we get,

$$p_\beta(m_0|h_0) = \int \prod_i dP(J_{i0})\ dm_i\ p_{\beta'}(m_i|h_i + k_i(m_i, m_0))\ \mathcal{K}(\mathbf{m}, m_0) \tag{23}$$

with,

$$\mathcal{K}(\mathbf{m}, m_0) = \delta\left[\tanh^{-1} m_0 - \beta J_{0j} m_j + \beta^2(1-q)m_0 - h_0\right]\ \exp[\beta^2(1-m_0^2)(1-q)]\ \frac{1}{1-m_0^2} \ , \tag{24}$$

and where we recall that, $\beta' = \beta[1 - 1/(2N)]$. The exponential in (24) is the factor ω, while the last term is the Jacobian of the delta. Let us compare equation (9) to (23): the key point is that in the general case there is no closed equation for $p(m_i)$, however complicated. On the other hand, if we start from a system with external magnetic field, the effect of the new spin can be counterbalanced by tuning the local fields in the new system, in such a way to leave stable the structure of states. For this reason it is possible to write a self-consistent equation only for the probability $p(m_i|h_i)$, but *not* for $p(m_i)$. This is the core of our approach.

In order to solve equation (23) we adopt an ansatz for the asymptotic form of $p(m|h)$ (we drop now the subscript 0). For small external fields, we expect,

$$p(m|h) = p(m)\ \exp\left[A(m)h^2 + B(m)h\right]\ , \tag{25}$$

where terms of higher order in h have been discarded, and $A(m)$ and $B(m)$ are functions to be determined self-consistently. By setting $h_i = 0$ and $h_0 \sim k_i \sim 1/\sqrt{N}$ in equation (23), and by using an integral representation for the delta function, we get to leading order in N:

$$\begin{aligned} p(m|h) \ &= \ C \exp\left\{-\frac{1}{2\beta^2 q}[\tanh^{-1} m - \Delta m]^2 + \lambda m^2\right\}\ \frac{1}{(1-m)^2}\ \times \\ &\qquad \times \exp\left\{-\frac{1}{2\beta^2 q}\,h^2 + \frac{1}{\beta^2 q}[\tanh^{-1} m - \Delta m]\,h\right\}\ . \end{aligned} \tag{26}$$

with $q = \langle m^2\rangle$, $A = \langle A(m)\rangle$, $\Delta = \beta^2\langle mB(m)\rangle - \beta^2(1-q)$ and $\lambda = (\beta^2/2)[\langle B(m)^2\rangle + A] + \Delta$. We note that the shift in temperature only gives a constant term at the leading order: together with all other constant contributions it has been reabsorbed in the normalization constant C. By comparing (25) with (26) we immediately have, $A(m) = -1/(2\beta^2 q)$, $B(m) = [\tanh^{-1} m - \Delta m]/(\beta^2 q)$. The parameters Δ and λ therefore satisfy the equations:

$$\begin{aligned} \Delta \ &= \ \frac{1}{2q}\langle m\ \tanh^{-1} m\rangle - \frac{\beta^2}{2}(1-q) \\ \lambda \ &= \ \Delta - \frac{1}{2q} + \frac{1}{2\beta^2 q^2}\langle[\tanh^{-1} m - \Delta m]^2\rangle\ . \end{aligned} \tag{27}$$

The expression of $p(m)$ we have found (first line in equation (26)), and the equations for the parameters Δ and λ are precisely those of the supersymmetry-breaking solution found long ago in [15]. The supersymmetric expression is recovered for $\Delta = \lambda = 0$ [7].

In this work we have shown how the cavity method can be generalized to situations where the structure of states is unstable under external perturbations, i.e. where the supersymmetry is broken. In this case, the simple distribution $p(m)$ is not stable when a new spin is added to the system. However, the perturbation caused by the new spin is analogous to an external field. Therefore, if we consider the system in presence of a field, we can balance the effect of the new spin by tuning the field. This means that the distribution $p(m|h)$ is stable when the new spin is added, and a self-consistent relation can be written for it. It is important to stress that this equation can be written for any system, once the form of the effective field k_i is known. In particular, applications of this method to diluted systems are under study. We thank F. Ricci Tersenghi for many interesting discussions.

[1] M.Mezard, G.Parisi and M.A. Virasoro, Europhys. Lett. **1**, 77 (1986).
[2] M.Mezard, G.Parisi and M.A. Virasoro, *Spin Glass Theory and beyond*, World Scientific, Singapore (1987).
[3] D.J. Thouless, P.W. Anderson and R.G. Palmer, Phil. Mag., **35**, 593 (1977).
[4] C. Becchi, R. Rouet and A. Stora, Comm. Math. Phys. **42**, 127 (1975); I.V. Tyutin *Lebedev preprint* FIAN 39 (1975).
[5] G. Parisi and N. Sourlas, Phys. Rev. Lett. **43**, 744 (1979).
[6] A. Cavagna, J.P. Garrahan and I. Giardina, J. Phys. A **32**, 711 (1998).
[7] A. Cavagna, I. Giardina, G. Parisi and M. Mezard, J. Phys. A **36**, 1175 (2003).
[8] A. Crisanti, L. Leuzzi, G. Parisi, and T. Rizzo, Phys. Rev. B **68**, 174401 (2003).
[9] T. Aspelmeier, A. J. Bray, M. A. Moore, Phys. Rev. Lett. **92**, 087203 (2004)
[10] A. Cavagna, I. Giardina and G. Parisi, Phys. Rev. Lett. **92**, 120603 (2004)
[11] G. Parisi and T. Rizzo, Preprint cond-mat/0401509.
[12] D. Sherrington and S. Kirkpatrick, Phys. Rev. Lett. **32**, 1792 (1975).
[13] T. Rizzo, Preprint cond-mat/0403261; Preprint cond-mat/0404729.
[14] This approximation is in fact exact in the present case, since we do not impose any constraint on the free energy.
[15] A.J. Bray and M.A. Moore, J. Phys. C **13**, L469 (1980).